

**A COMPARATIVE STUDY ABOUT COGNITIVE LOAD
OF AIR GESTURES AND SCREEN GESTURES FOR
PERFORMING IN-CAR MUSIC SELECTION TASK**

A Thesis
Presented to
The Academic Faculty

by

Xiaolong Wu

In Partial Fulfillment
of the Requirements for the Degree
Master of Industrial Design in the
School of Industrial Design

Georgia Institute of Technology
December 2014

Copyright © 2014 by Xiaolong Wu

**A COMPARATIVE STUDY ABOUT COGNITIVE LOAD
OF AIR GESTURES AND SCREEN GESTURES FOR
PERFORMING IN-CAR MUSIC SELECTION TASK**

Approved by:

Dr. Young Mi Choi, Advisor
School of Industrial Design
Georgia Institute of Technology

Dr. Bruce Walker
School of Psychology and School of
Interactive Computing
Georgia Institute of Technology

Prof. Tim Purdy
School of Industrial Design
Georgia Institute of Technology

Date Approved: 12/04/2014

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Young Mi Choi, not only for being a great advisor of my thesis, but also for all her support throughout my graduate career. I am grateful for it. I wish to thank Dr. Bruce Walker and the Sonification Lab for their generous help during my thesis development. I appreciate it a lot. I want to thank Prof. Tim Purdy for being a great mentor in my thesis defense, as well as being a great professor in the 3D modeling and rendering class. I also want to thank all the participants for completing my user studies, my families and friends for their help in my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
SUMMARY	ix
I INTRODUCTION	1
1.1 Significance of the Problem	1
1.2 Goals of the Study	2
II BACKGROUND	3
2.1 Visual Distraction and Cognitive Load	3
2.2 Gesture Modality vs Speech Modality	4
2.3 Gesture Recognition Techniques	6
2.4 Feedback	8
2.5 Simulator Fidelity	9
2.6 Gesture Types	10
2.7 Screen Gesture vs Tactile Button	10
2.8 Air Gesture vs Tactile Button	14
2.9 Hypothesis	14
III METHODOLOGY	16
3.1 Design	16
3.2 Procedure	24
IV RESULTS	26
4.1 Secondary Task Completion Performance	26
4.2 Subjective Workload	26
4.3 User Preferences	27
4.4 Driving Performance	29

4.5	Eye Glance Behavior	30
V	DISCUSSION	32
5.1	Secondary Task Completion Performance	32
5.2	Subjective Workload	33
5.3	User Preferences	33
5.4	Driving Performance	34
5.5	Eye Glance Behavior	34
5.6	Limitation	35
VI	CONCLUSION	37
APPENDIX A	— SECONDARY TASK PERFORMANCE DATA	38
APPENDIX B	— NASA TLX WORKLOAD DATA	47
APPENDIX C	— USER PREFERENCES DATA	49
APPENDIX D	— DRIVING PERFORMANCE DATA	53
APPENDIX E	— EYE GLANCE BEHAVIOR DATA	55
REFERENCES	58

LIST OF TABLES

1	Median value of each interface with regard to the four aspects (N = 14).	29
2	Secondary Task Completion Time Raw Data(in milliseconds) (subject 01).	39
3	Secondary Task Completion Time Raw Data(in milliseconds) (subject 02).	39
4	Secondary Task Completion Time Raw Data(in milliseconds) (subject 03).	40
5	Secondary Task Completion Time Raw Data(in milliseconds) (subject 04).	40
6	Secondary Task Completion Time Raw Data(in milliseconds) (subject 05).	41
7	Secondary Task Completion Time Raw Data(in milliseconds) (subject 06).	41
8	Secondary Task Completion Time Raw Data(in milliseconds) (subject 07).	42
9	Secondary Task Completion Time Raw Data(in milliseconds) (subject 08).	42
10	Secondary Task Completion Time Raw Data(in milliseconds) (subject 09).	43
11	Secondary Task Completion Time Raw Data(in milliseconds) (subject 10).	43
12	Secondary Task Completion Time Raw Data(in milliseconds) (subject 11).	44
13	Secondary Task Completion Time Raw Data(in milliseconds) (subject 12).	44
14	Secondary Task Completion Time Raw Data(in milliseconds) (subject 13).	45
15	Secondary Task Completion Time Raw Data(in milliseconds) (subject 14).	45
16	Average Secondary Task Completion Time(in milliseconds), N = 14.	46
17	Number of Secondary Task Errors(in times),N = 14.	46
18	Overall Workload Score(in points), N = 14.	48

19	Overall Preference (N = 14).	51
20	Effectiveness(N = 14).	51
21	Efficiency(N = 14).	52
22	Satisfaction(N = 14).	52
23	Average Brake Response Time (in milliseconds), N = 14.	54
24	Average Longitudinal Variance (in meters), N = 14.	54
25	Total Eye Dwell Duration (in milliseconds), N = 14.	56
26	Total Number of Dwells (in times), N=14.	56
27	Total Number of Long Dwells (in times) ,N = 14.	57
28	The Dwell Rate Per Task, N = 14.	57

LIST OF FIGURES

1	How a hand tracking device works with infrared lights and depth cameras.	7
2	The typical tactile buttons and knobs for in-car secondary tasks. . . .	11
3	Tesla is using a multi-touch screen as its informative display.	12
4	Participants use air swipe gesture to change soundtrack. The image on the left means 'Previous', the one on the right means 'Next'. . . .	17
5	Participants use air tap gesture to change soundtrack. If the direction of the index finger is left, it indicates 'Previous', while it is right, then it indicates 'Next'.	17
6	A Leap Motion device with a USB cable.	18
7	The Top view of the study layout.	19
8	The participant is tapping on the screen button to change soundtrack.	20
9	The image on the left shows the current soundtrack name is '5 o'clock', and the current task name is 'All I want for Christmas is you' while the one on the right shows the task is completed.	21
10	One subject is performing the driving task on the low fidelity simulator	23
11	The image on the left shows the location and the appearance of the facelab 4 eye tracker, while the one on the right shows a picture in which one subject's pupils are being tracked.	24
12	Average Task Completion Time(in milliseconds) N=14.	27
13	Subjective Workload, N=14	28
14	Average Brake Response Time (in milliseconds) N = 14.	30
15	Average Longitudinal Variance (in meters) N = 14.	31

SUMMARY

Since gestures are intuitive and relatively contact free, they have been considered as a feasible way for performing in-car secondary tasks. However, few researches have been conducted in terms of subjects' cognitive load. This thesis examined four gesture interfaces (air swipe, air tap, screen swipe, and screen tap), in terms of their effects on drivers' driving performance, secondary task performance, perceived cognitive load, and eye glance behavior. The result demonstrated that air gestures are generally slower than screen gestures with regard to secondary performance. Screen swipe gesture requires the lowest cognitive load while air swipe and screen tap gesture remain the same. Subjects in this study tend to prefer screen swipe gesture the most while prefer air tap gesture the least. However, there is no significant difference between air swipe and screen tap gesture. Although air tap gesture and screen tap gesture generated the largest amount of dwell times, no variance among the four gesture interfaces in driving performance has been found. The result indicated that even though air gestures are not limited by space, screen swipe in this study still seemed to be the most ideal way for performing in-car secondary task of music selection.

CHAPTER I

INTRODUCTION

With the development of technology, people's viewpoints of the automobile have shifted; instead of merely a means of transportation, the automobile has become a space in which a driver can still perform daily activities besides driving, such as communicating with other people, interacting with electronic devices, and receiving information. In the meantime, different ways of interaction have been explored. Among all the modalities, speech and gesture have gained the most popularity because of their intuitiveness. In order to implement a modality control for in-car use, safety is the most important factor the car manufacturer should pay attention to. This paper will focus on comparing the cognitive load and visual distraction of different types of gesture interaction on drivers.

1.1 Significance of the Problem

Taking phone calls, communicating with passengers, switching music channels, or controlling in-car temperature are some of the common secondary tasks that drivers now perform. Also, these secondary tasks provide the drivers a better driving experience. However, because of the limitation of humans' capacity for processing multiple tasks [13], interacting with in-car systems while driving will easily lead to accidents. Dingus .et.al [16] found that almost 80% of car crashes involved driving distraction 3 seconds before the accident. Moreover, among the in-car information and entertainment systems, music control tends to appear in the top six distractors for drivers [17]. Researches have found that one main driving distraction is losing visual attention, and the other one is having great cognitive workload [12]. Since safety is regarded as the primary concern of driving, before a new way of interaction is developed and

implemented, the effects of the interaction techniques should be carefully examined.

1.2 Goals of the Study

Previous research studies focused more on the effects of gesture control on drivers' driving performance. Gestures and other input modalities have been evaluated in terms of driving, secondary task performance and visual distraction. However, few articles addressed the field of comparing cognitive load, which is also very important in the driving activity. On the other hand, even though gesture seems to be a proper technique for in-car secondary task because of its contact free and intuitiveness, the impacts of different types of gestures may vary. Therefore, a comparative study has been conducted in this paper, to evaluate the cognitive load, visual attention, driving and secondary task performance that touch and air gestures have.

CHAPTER II

BACKGROUND

In this section, the impact of visual distraction and cognitive load are presented. In order to demonstrate the reason why gesture control is chosen in this paper, the pros and cons of gesture control, the difference between different types of gestures, and the current gesture recognition techniques are also illustrated. The current comparative research studies between gesture control and tactile interface are also described and discussed.

2.1 Visual Distraction and Cognitive Load

Visual distraction refers to the driver's behavior of diverting his or her vision from the road in order to help locate or manipulate other devices in car [2]. Losing visual attention on the road will lead to the delay of lane tracking response time and the temporary fixing of steering wheel angle [12]. Since the angle of steering is not able to change based on the road condition, studies have found that visual distraction has significant effects on drivers' lane keeping performance. Poor lane keeping performance will easily lead to lane weaving and lane exiting [25].

On the other hand, losing visual attention also results in speed reduction[12]. In Antin [27]'s study, researchers compared subjects' driving performance between reading a static map and reading an electronic map. They found that the increasing of drivers' gaze times has led to speed reduction. One explanation is that when drivers lose visual focus on the road, they try to reduce their speed to maintain their lane keeping performance. Driving while performing secondary tasks makes it easy for drivers to lose visual attention, because some interaction techniques require users' visual attention. Campbell, et al [26] have compared the visual distractions that

different types of displays create, and they found that the complexity of the visual display significantly affected drivers' steering wheel reverse rate. (Importance)

Another big factor that affects driving performance is high cognitive load. In-car cognitive load tasks refer to the tasks that require no visual switch between the road ahead and the task itself [12], such as talking on a hand free phone, or using voice command. High cognitive load has been found to increase drivers' reaction time. By conducting user studies requiring subjects to have conversations on the phone while driving, Horrey.et.al [28] have found that cognitive load does not affect the lane-keeping performance, while surprisingly increases their driving speed. However, their reaction time became significantly longer. The more complex the conversation is, the longer the reaction time became. On the other hand, high cognitive load also leads to more steering wheel correcting activity times [12].

Previous studies [1] [5] [6] [7] mainly focused on the visual distraction aspect of modalities' input. For cognitive load, they only examined simple and traditional secondary tasks, such as communicating with other people, or interacting with electronic devices. As the number of in-car secondary task increases, so as the number of ways of performing these tasks, more work needs to be done to evaluate the new ways of interaction.

2.2 Gesture Modality vs Speech Modality

In order to provide drivers a pleasant user experience while keeping the driving activity safe, automobile manufacturers and researchers are exploring different modalities for performing secondary tasks. The main goal of this exploration is to find modalities that minimize the visual distraction and the cognitive load while maintaining the driving performance. Apart from traditional tactile buttons, the two modalities that have gained the most popularity are gesture and speech.

Speech input, which does not require visual load and is hands-free, has been

evaluated in previous researches for in-car use. For instance, when the driver wants to change the volume, instead of finding the button or knob, he/she can just command 'volume up'.

However, a study result has shown that using speech input while driving required significantly high cognitive load, especially when the system became more complex. Moreover, by comparing subjects' driving performance between one with a speech system and one without, the former has been found to impair the driving performance by increasing drivers' braking response time [14]. The study also found that the complexity of the speech system did not affect drivers' reaction time. On the other hand, the amount of cognitive load that the speech input requires varies among tasks. A study by the AAA Foundation for Traffic Safety [22] has examined several current speech recognition systems by a five-category rating criteria. The result showed that composing text messages and emails require higher cognitive load than listening to messages.

Moreover, it's difficult for users to give valid instructions to a speech system. For example, it's not natural for the drivers to say, 'Raise the music volume 20% higher'. The last but most important problem is, speech input can't be conducted in a noisy environment. The automobile is a space where there is lots of noise, such as the sound of the engine and wheel, the in-car music, and passengers' conversations. All this noise will to some extent reduce the performance and accuracy of the speech recognition system. This is important because the accuracy of the recognition system has been proven to affect drivers' cognitive load significantly [22]. Therefore, speech recognition for in-car use has been regarded as an impractical solution [32].

In-car gesture control refers to the movement that drivers perform with one of their hands to conduct a series of commands while driving. Though gesture control is limited in command vocabulary size, and some types of gestures require cognitive load from users [2], Akyol.et.al [4] have found that subjects regarded the gesture

interaction as natural and intuitive after they tested static gestures on a computer-vision based interface. Furthermore, since gesture control does not require high visual load, it is considered as a feasible way for performing secondary in-car tasks [2]. In the meantime, as the number of secondary tasks increases and the in-car space remains limited, gesture seems to be a potential replacement or complement for traditional buttons [4]. Nowadays, automobile manufacturers, such as Audi, BMW, Ford and Mercedes-Benz are all working on developing gesture control system for their vehicles [10].

2.3 Gesture Recognition Techniques

Currently, there are mainly two techniques for gesture recognition. Computer vision technology has been widely explored and implemented in gesture recognition area. The traditional gesture recognition process is based on IR cameras and infrared LEDs. The LEDs generate 3D patterns of users' hand gesture by LED dots, and then the camera will capture the pattern and send it to the host computer for analysis [8] (Figure 1). For in car gesture recognition, the advantage of computer vision technology lies in its non-intrusiveness because it captures the gesture remotely. The recent popular gesture or body tracking devices, such as Kinect and Leap Motion [8], both apply this technology. However, when applying the computer vision technology to in-car use, there is a problem when the automobile is moving in the daylight, the sunlight will largely affect the infrared lights; Therefore, the IR camera will not be able to capture the gestures in the real car-moving environment. To solve this problem, Lyons [19] developed a gesture library that used machine learning to pre-train the system, and then used a RGB camera to receive the shadow of the infrared light. However, in this study, since leap motion is simple, it is suitable for proof of concept. This paper will concentrate on computer vision based gesture control.

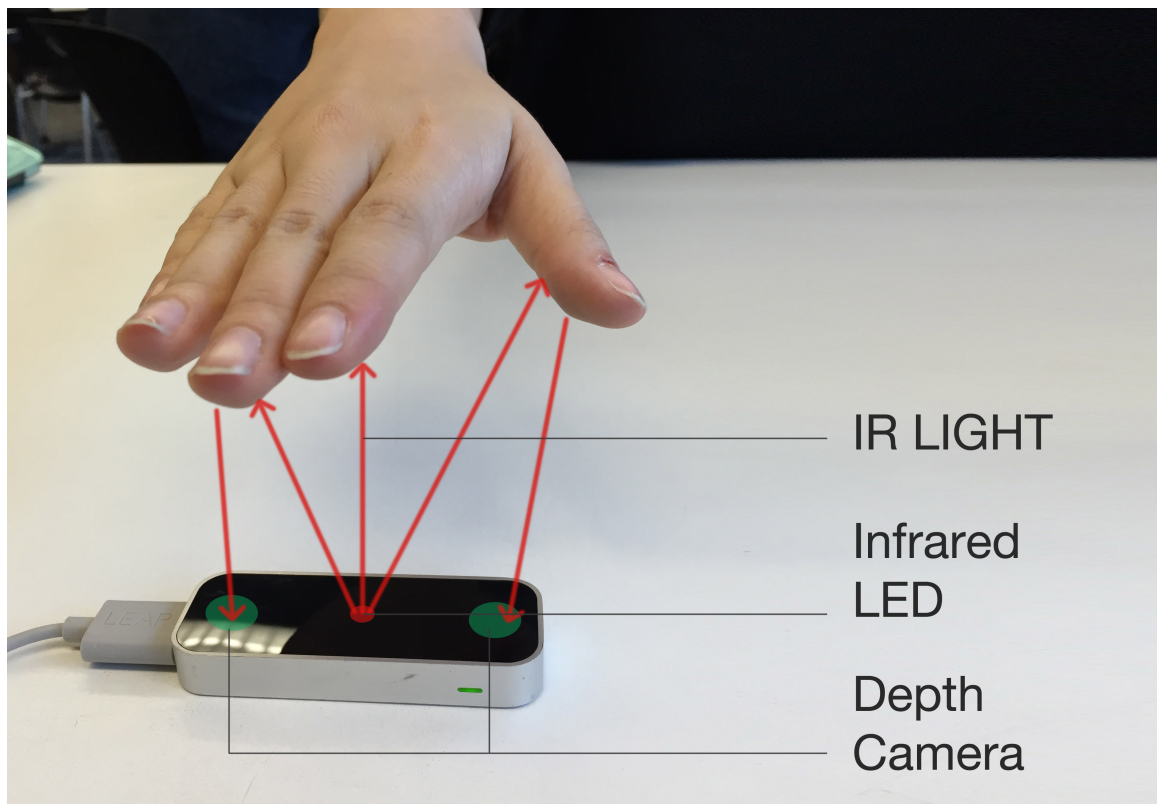


Figure 1: How a hand tracking device works with infrared lights and depth cameras.

The other technique is sensor-based technology, which requires users to wear additional devices, such as a glove. Gyroscope and accelerometer sensors are always used to measure the motion of the user's hand. The advantage of sensor-based technology for in-car use is that drivers are able to issue commands by subtle finger movements, that the hand does not need to leave the steering wheel [15], and that it could provide direct physical feedback, such as vibration, without being interfered by sunlight and environment. However, since it requires users to wear additional devices on the hand, some gloves also connect with the computer via several wires, users think it's inconvenient and uncomfortable [15]. Therefore, sensor based gesture recognition is not discussed in this paper.

2.4 Feedback

Feedback is to inform the user of the current context, what action has been done and what has been accomplished. Therefore, it is essential in the process of interaction. There are basically three types of feedback for in-car use. The first one is traditional tactile tangible feedback. When a user presses a button or rotates a knob, he naturally receives the tactile feedback. It is either a rebound of a button or the friction of the gear. Another type of feedback is auditory feedback. It can tell the user the current state of a task and it does not require any visual attention. The third one is visual feedback. The current state of the task will be shown on the informative display, either by text or by image.

Since drivers have to switch their visual focus between the front road and the head unit in terms of visual feedback while they do not need with audio, in order to evaluate the effects of visual and audio feedback on drivers, Christiansen .et. al [1] has conducted a study comparing audio and visual feedback in terms of driving performance, secondary performance, and eye glance behavior. The result showed that visual feedback generated fewer number of speed increase times than audio feedback,

while there was no significant speed deviation. One contributor that Christiansen believed to the more speed increase times is that audio feedback required subjects to listen and then process the task, thus it required higher cognitive load and led to poorer driving performance.

In terms of secondary task performance, visual feedback generated more errors, and the task completion time was significantly faster than audio feedback. For the number of eye glance times, the result indicated that audio feedback generated significantly less glances than visual feedback.

Therefore, in order to provide users with feedback and decrease the visual distraction, in this paper, auditory feedback, along with an informative display will be provided in the user study.

2.5 Simulator Fidelity

A typical driving simulator consists of a steering wheel, a brake pedal and an accelerator that simulates the real driving activity. Several screen displays are used to simulate the views that the user sees in the driving environment. Studies have indicated that the driving performance on the driving simulator is predictive for real on-road driving performance [24]. Furthermore, since it is safe, it is easy to measure subjects' driving performance, and it is able to simulate various driving conditions [23], most of the current driving related researches have used driving simulators for user studies.

Jaegar et al [3] has compared the driving performance and secondary task performance between the real controlled driving situation and the simulated one. The result showed that there was no variance for secondary tasks. However, controlled driving situation was found to generate more long (2.0 seconds) eye glances. Moreover, subjects' longitudinal performance was significantly better in controlled driving situation than simulated one.

2.6 Gesture Types

Gestures are regarded as a natural way in which human beings express ideas [2]. However, there are many different types of gestures and their effects on users vary greatly. Thus, the overall feasibility of implementing gestures for in-car use, as well as the types of gestures that are suitable have been carefully evaluated in previous studies. There are mainly three types of gesture, natural hand gesture, symbolic gesture and sign language gesture. Natural hand gesture refers to a set of gestures that respond to natural hand motion, such as changing hand direction and position. Since it does not generate high learning curve, and requires the lowest cognitive load from users, it is ideal for in-car use. However, the vocabulary size for natural hand gesture is limited. It can only be applied to control a few simple features that have a clear argument, such as volume or temperature. Symbolic gesture, a type of gesture that is always used to simulate the physical environment (swipe, tap, wheel), is also used for in-car interaction. Even though users have to pre-memorize a certain set of commands, it is natural and intuitive. The command vocabulary size is also larger than that of natural hand gestures. Sign language is useful when it is applied to translating from gestures to low level continuous speech [2]. Compared to the two gesture types above, it is more powerful. However, sign language gesture has been found to require significant training time (100 hours of practice)[29]. Therefore, sign language is not considered to be suitable for in-car use. In this paper, simple symbolic gestures that control a specific set of task is discussed.

2.7 Screen Gesture vs Tactile Button

Nowadays, touch screen interfaces have been widely used in cars. As technology advances, the number and the complexity of secondary tasks that can be conducted in the car also increases. Given the limited space in the automobile, touch screens provides more flexibility than traditional physical buttons by synthesizing several tasks



Figure 2: The typical tactile buttons and knobs for in-car secondary tasks.



Figure 3: Tesla is using a multi-touch screen as its informative display.

on a single screen. Furthermore, compared to physical buttons, the cost is relatively lower [1]. On button-based touch screens, physical buttons are replaced by icons of buttons on the screen. Users make commands by pressing buttons on the screen. However, though it saves space, compared with traditional tactile button interface (shown in Figure 2), the touch screen interface has been found to be more mental and visual demanding and generates poorer driving performance (lateral deviation) [31] [33]. Noy et al. [30] also found that since the virtual buttons on touch screen lacks tactile feedback, users tend to use their vision to target at it. Hence, it requires significant visual load.

Therefore, on the touch screen, besides traditional virtual touch button, gesture-based interaction is gaining popularity in performing secondary tasks (Figure 4). Users use simple gestures, such as directional swipe, to make a command on a multi-touch screen. In Christiansen. et. al[1]’s study, screen-based gesture interaction did not improve users’ driving and secondary task performance, and it also generated

more driving errors in the study. However, in terms of eye glances, they examined the number of total, quick (below .5s), medium (.5s - 2s), and long (above 2s) glances. The reason why they categorize glance is that they believed that it took less than .5s for subjects to locate the screen when they interacted with touch screen gesture. They hypothesized gesture would generated more quick glances while touch buttons generated more medium and long glances. However, it has been found that touch button generated 51% more quick glances; almost twice more medium glances than touch screen gesture. Furthermore, the combination of screen gesture and audio feedback was found to significantly reduce the number of eye glances.

Similarly, Jaeger. et. Al [3] have also conducted a comparative study among traditional tactile buttons, virtual touch buttons, and touch gesture (simple directional gesture). They found that although traditional tactile interface generated significantly more lateral errors (lane exiting, steering wheel input) in driving task and caused the most number of eye glances, the study indicated that none of them generated significantly better longitudinal driving performance. The gesture-based interface required the least amount of visual load (least number of eye glances), while the virtual touch interface was the most efficient for completing secondary tasks. Since cognitive load was not measured, we predict that using gestures requires high cognitive load from the drivers, which reduced the driving performance.

Given the result that gesture-based touch screens generated less visual distraction, Tuomo [6] [7] has conducted several research studies comparing kinect scrolling, swipe, and touch button screen for the secondary task of music selection. The result showed that kinect scrolling generated the poorest driving performance and it also increased the visual demand. Swipe gesture has been proven to be more suitable for in-car use because it is intuitive and does not require a high degree of accuracy.

2.8 Air Gesture vs Tactile Button

As computer vision technology improves, air gestures have been introduced as a new way of in-car interaction. Compared with screen gestures, as there is no physical interface required at all [2], air gestures are not limited to a fixed space. Users issue commands by moving their hands in the air. Although some study results have shown that users tended to prefer simple directional air gesture interfaces to tactile buttons as it requires less visual input and users don't need to touch or reach anything, Alpern [5] has found that there is no significant driving performance difference between air gestures and traditional tactile buttons. However, since the secondary task performance and cognitive load were not measured in the study. We believed that the reason for the lack of variation in driving performance is due to the fact that while traditional buttons result in high visual load, air gestures require cognitive load, which both impairs the driving performance.

As for which air gesture is suitable for in-car use, even though no comparative studies have been conducted to find the appropriate mapping between tasks and different air gestures, a theater theory based study has been conducted [35], in which users' subjective preference have been measured. Users in the study were asked to map the functions and the gestures they would like to use. It found that users ranked swiping air gesture as the top one for performing incremental movement function.

2.9 Hypothesis

Studies have shown that screen gestures generated fewer eye glances than touch buttons. In the meantime, compared to other screen gestures, screen swipe gesture has been proven as the most efficient way for performing music selection task. However, screen gestures did not yield better driving performance than traditional tactile buttons and screen buttons. Therefore, we predicted that using screen gestures required

high cognitive load from the drivers. On the other hand, though few studies have addressed their cognitive load, air gestures were considered intuitive and natural by the subjects. Furthermore, air swipe gesture was also preferred the most for incremental movement over other gestures.

Based on previous research, we hypothesized that air gestures were more time consuming for music selection task, however, since users did not need to gaze at the visual display, fewer dwells would be generated compared to screen gestures. Additionally, air gestures did not limit subjects's hand on a certain screen, so we expect air gestures would generate less cognitive workload and lead to better driving performance. A study was conducted which compared air swipe gestures, air tap gestures, screen swipe gestures and screen touch gestures in terms of driving performance, eye glance behavior, secondary task performance and cognitive load.

CHAPTER III

METHODOLOGY

A pilot study with two subjects was conducted prior to the real user study. There were 14 participants recruited (Age, Mean = 24.81, SD = 1.50) in the real study. The inclusion criteria required all the subjects to have valid drivers' licenses, and to be in a fair health condition. None of the 14 subjects have the experience of playing with driving simulator before. 3 subjects claimed that they have used 'leap motion' before this study.

3.1 Design

Four types of interfaces, including air swipe (AS), air tap (AT), touch screen swipe (SS), touch screen tap (ST), have been compared in the study. For AS (shown in Figure 4), swiping towards the right means 'Next' while swiping towards the left stands for 'Previous'.

For AT gesture, users use their forefingers to tap forward in the air (shown in Figure 5). The lateral direction of their forefingers determines the action of the secondary task. Tapping to the right stands for 'Next' while tapping to the left means 'Previous'.

For both air gestures, the recognition process was carried out by 'Leap motion' [8]. The image of Leap Motion is shown in Figure1. It is a tiny device that utilizes infrared light and depth cameras for gesture recognition. Since the Leap Motion is able to detect hand gestures around 3 feet hemispherical area, it is powerful enough for in-car gesture control. Leap motion is connected with a laptop via an USB port. When turned on, leap motion keeps sending the captured image frames to the computer for analysis.

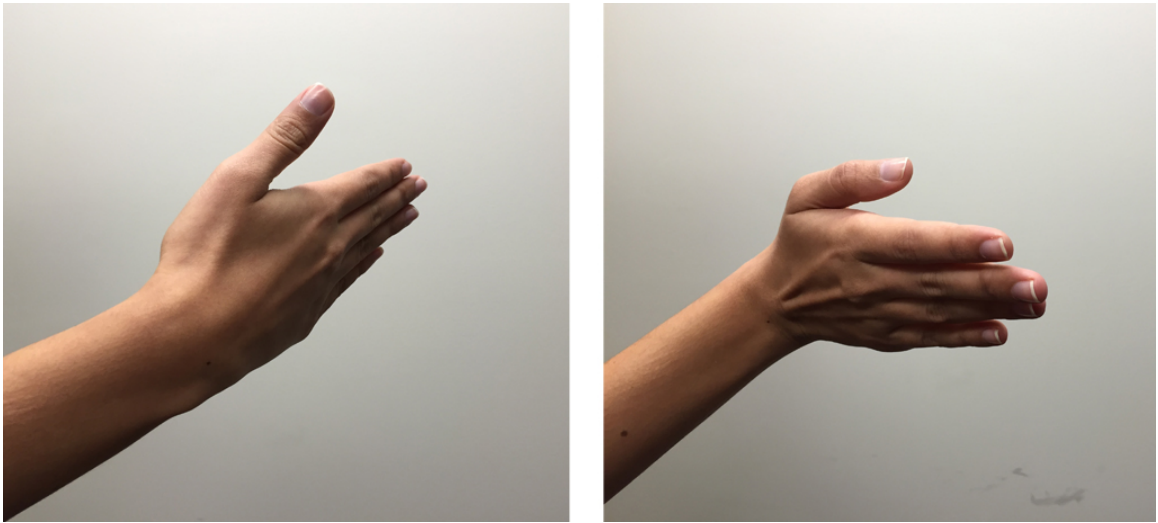


Figure 4: Participants use air swipe gesture to change soundtrack. The image on the left means 'Previous', the one on the right means 'Next'.

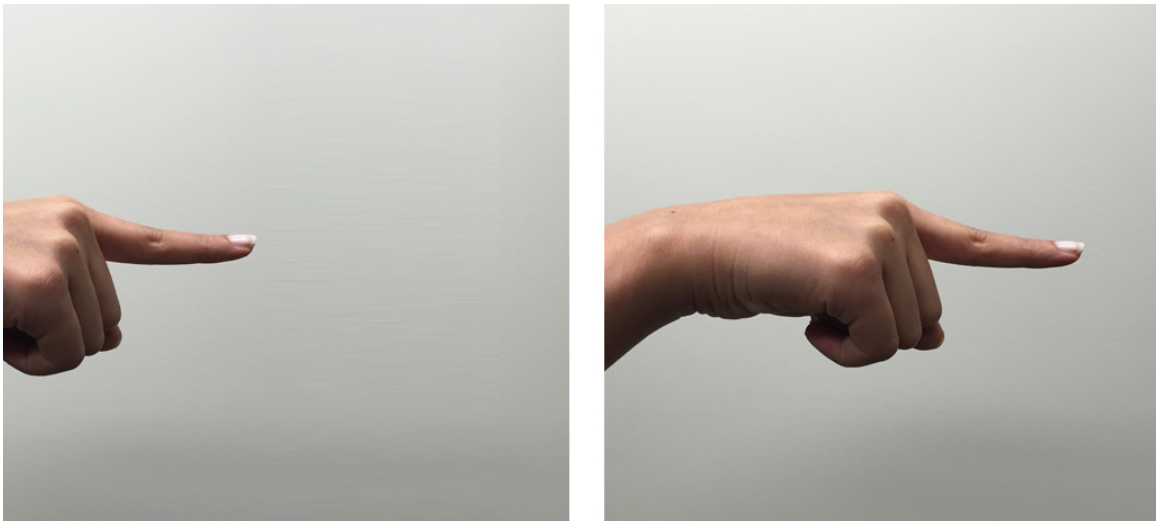


Figure 5: Participants use air tap gesture to change soundtrack. If the direction of the index finger is left, it indicates 'Previous', while it is right, then it indicates 'Next'.



Figure 6: A Leap Motion device with a USB cable.

Since leap motion is able to track hand gestures within 3 feet hemispherical area, and driver's seat is on the left side in U.S, the hand-tracking device was placed on the right side beside the subjects(shown in Figure 6). Moreover, it was located in front of the informative display. The exact location of the leap motion is fixed by several blue tapes during the whole user study.

For the two touch-based interfaces, a multi-touch tablet (shown in Figure 2), was used in this study. Since the in air gesture recognition was powered by a laptop and the touch interface was based on a tablet, in order to keep the consistency, a tablet application called air display was used to mirror the laptop screen on the tablet.

The graphical user interface shown in Figure 7 was used for providing users both visual and auditory feedback about the current soundtrack. There were two green buttons on bottom left and bottom right, allowing users to touch to change soundtrack. The current selection task appeared in the bottom center. The whole screen also allowed users to perform touch swipe gesture. The secondary task program was a web application coded in Javascript. There was a library that consists of names from

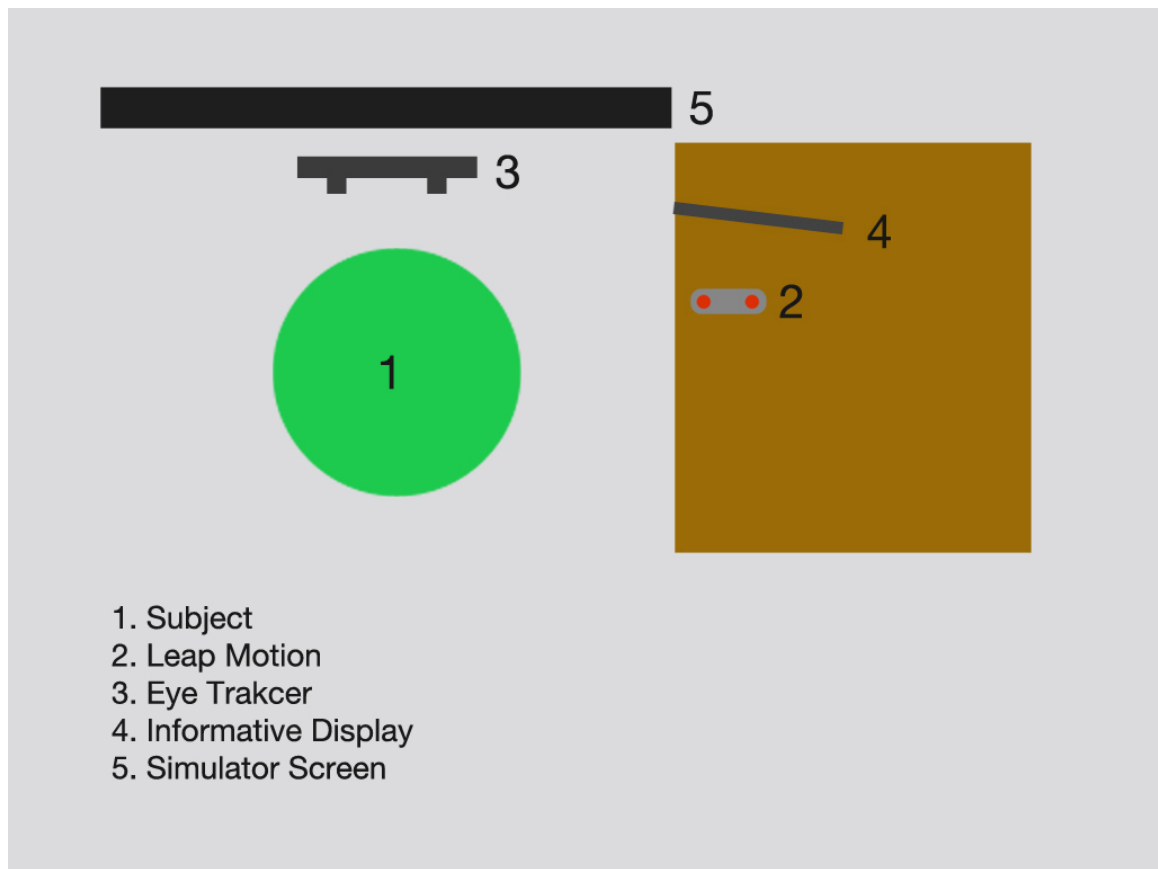


Figure 7: The Top view of the study layout.



Figure 8: The participant is tapping on the screen button to change soundtrack.

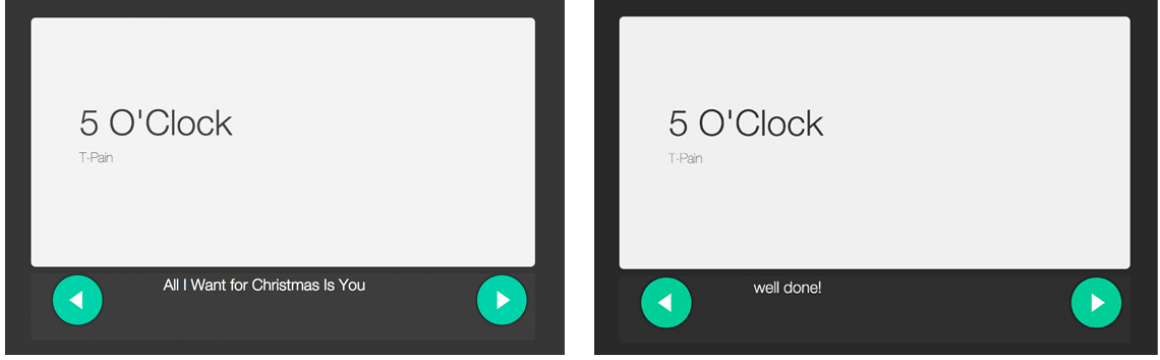


Figure 9: The image on the left shows the current soundtrack name is '5 o'clock', and the current task name is 'All I want for Christmas is you' while the one on the right shows the task is completed.

50 different soundtracks. They were sorted alphabetically. In order to measure subjects' secondary task performance, a set of 14 music selection tasks was given to the subjects while they were driving on the simulator in each condition. The order of the first 2 tasks was fixed, so as to flatten the learning curve. The order of the rest tasks was randomized every time when assigned to the subjects. Each music selection task required the subjects to find a certain soundtrack in the program. The difficulty of each music selection task ranged from level 1 to 20, whereas level 1 indicated finding the first song in the library. Each task was triggered by the experimenter, announced by a text-to-speech system. A spindex system [9] was used to provide the subjects audio feedback (soundtrack name) when the subjects browse in the library. When the auditory task command was assigned, the task was also shown in bottom center on the informative display (shown in Figure 8).

In the real driving environment, drivers do not need to press 'play' button every time they change the soundtrack. In order to simulate the real scenario, there was no extra action needed to confirm the music selection task. Once the subjects felt that they have completed a given task, they can move their hand back on the steering wheel. If the program did not sense any air gesture or touch in 8 seconds, the task was considered to be complete.

The software recorded the completion time of each task by subtracting 8 seconds. The duration was set to be 8 seconds because in some driving situations (eg. the lead car brakes, the following car changes lane), subjects tend to move their hands back to the steering wheel to maintain driving performance, regardless of if the secondary task was completed. Therefore, 8 seconds provided the subjects with enough time to handle driving task and continue secondary task afterwards.

The driving simulator in the study was a low fidelity simulator. (shown in Figure 9). It contained a screen display that simulated the front view, a steering wheel, an accelerator and a braking pedal. The manipulation of the simulator was the same as the real driving activity. There was a blue car in front of the subject's car, and there was a red car behind. The subjects were told to keep a safe distance (around 20 meters) between the lead car. The lead car braked randomly during driving. Subjects should respond to it by stepping on the brake pedal as quickly as possible. The car behind randomly turned on its left turn signal, which the subjects can see through the rear-view mirror on the top of the screen. As soon as they noticed it, they should press a button on the steering wheel to let the system know that they have noticed it. To measure subjects' eyeball movements, a facelab4 eye tracker (shown in Figure 10) was used in the study. It sit in front of the subject, and measured the subjects' eyeball movement silently. Before each user began the study, the eye tracker monitored the eyes of the user, and then it built a model. The eye tracker was able to tell whether the user was looking at the front view or the informative display while he was performing dual tasks. All the eyeball movement data was stored in the computer that it connected with for analysis.

Four aspects of eyeball movements were examined in this study, including total dwell duration, the number of dwells, and dwell rate per task the long dwell rate per task (above 1.6s). Duration of 1.6s was regarded as long dwell because study [3] has suggested that drivers would feel uncomfortable without looking at the road for such



Figure 10: One subject is performing the driving task on the low fidelity simulator a period of time. Total dwell duration and the number of dwells were used in order to compare the visual output of the four gestures. With regard to the dwell rate per task, it was used in order to measure how much time subjects spent for completing one task.

The NASA TLX survey [34] was used in this study to measure subjects' perceived workload about each interface. The survey examined six aspects of a task, including mental demand, physical demand, temporal demand, performance, effort, and frustration, by letting the subjects compare each pair of aspects based on their subjective opinions. The output of the survey was a score ranging from 0 to 100; it indicated the overall workload score. The higher the score was, the higher the workload the subject perceived.

Besides the NASA TLX survey, a questionnaire was also used to measure subjects' subjective preferences. Subjects in the questionnaire were asked to rank the four types

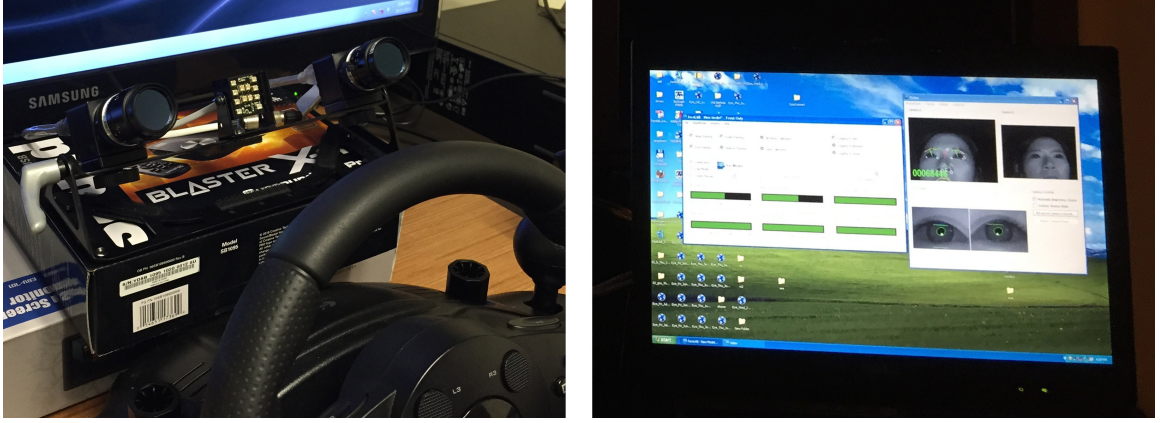


Figure 11: The image on the left shows the location and the appearance of the facelab 4 eye tracker, while the one on the right shows a picture in which one subject's pupils are being tracked.

of interfaces based on their perceived overall preference, effectiveness, efficiency, and satisfaction.

3.2 Procedure

The experiment procedure was described as following. At first, the participants were given an instruction about the goal of the research and an introduction of the four interfaces (AS, AT, SS, and ST). Then, the subjects were given a tutorial about the driving simulator, including their primary driving tasks. While the subjects were practicing with the driving simulator, their eyes were tracked and modeled by the remote eye tracker. After the subjects stated to the experimenter that they were comfortable with the driving simulator, they would be given another introduction to the informative display, which was used for performing the secondary task.

There were five conditions in the study, including a single driving task (Baseline) as well as four dual driving tasks with each of the interfaces (AS, AT, SS, and ST). In order to minimize the order effects, the sequence of the five conditions was generated by using a Latin Square equation. Before the experiment began, all the participants were informed that the driving task was the most important one, and they could perform the secondary task at their own pace.

There were two sessions for each condition. All the subjects were given a short practice session (8 subtasks) before the test session started. After the practice session, the participants would start the test session. Each of the secondary tasks was triggered by the experimenter. The first subtask was assigned to the subject by the experimenter after his or her car passed the starting sign in the simulated environment. There were 14 subtasks in total for each condition. A break of a few seconds was given between each subtask. After each condition was completed (other than Baseline Driving), the subjects were asked to fill in a NASA TLX Survey. After subjects completed all five conditions, they were asked to answer a questionnaire with regard to their personal preferences.

CHAPTER IV

RESULTS

4.1 Secondary Task Completion Performance

The completion time of each music selection task was recorded by the application itself. There were 14 tasks in each test session. Since the first 2 tasks were still regarded as training session, they were subtracted before the analysis. The total tasks for each condition were the same, but the difficulty of each music selection task varied. In order to compare the task completion time of each condition, the mean value of each condition for each subject was calculated first. Then it was analyzed by a within subjects ANOVA test (two tailed, confidence interval adjusted by Bonferroni). The result showed that there was a significant difference between screen gestures and air gestures, $F(3,39) = 19.882$, $p < .05$. Both SS ($M = 11.12s$, $SD = 2.71$) and ST ($M = 11.44s$, $SD = 1.86$) were significantly faster than AS ($M = 17.33s$, $SD = 3.42$, $p < .001$) and AT ($M = 20.46s$, $SD = 6.69$, $p = .002$). However, there was no significant difference between SS and ST, and between AS and AT.

For the number of task errors, the average number of task errors of each condition for each subject was collected by the application. The data was analyzed by a repeated measures within subjects ANOVA test (two tailed, confidence interval adjusted by Bonferroni). The result indicated that there was no significant difference among the four conditions, $F(3,39) = 2.6$, $p > .05$.

4.2 Subjective Workload

After subject finished each condition, they were asked to fill in a NASA TLX survey. It was used to measure the perceived workload of each condition. The overall task

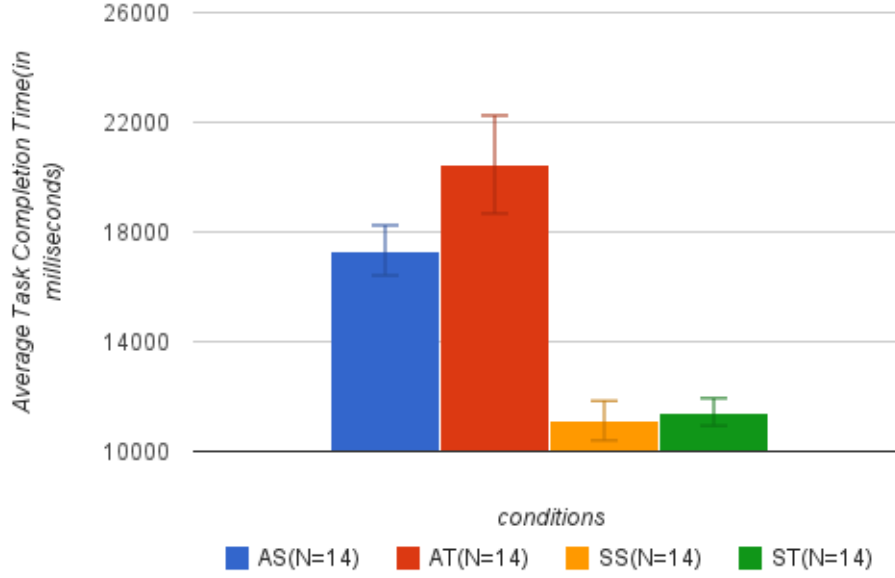


Figure 12: Average Task Completion Time(in milliseconds) N=14.

load of each gesture interface was presented in Figure 13. The data was analyzed by a within subjects ANOVA test (two tailed, confidence interval adjusted by Bonferroni). Each condition within each subject was compared first, and then each condition for all subjects was compared. The result showed that there were statistical differences among the four interfaces, $F(3, 39) = 15.339$, $p < .05$. The workload of SS ($M = 33.09$, $SD = 4.43$) was significantly lower than that of AS ($M = 51.44$, $SD = 3.47$, $p = .028$), and AT ($M = 67.22$, $SD = 3.59$, $p < .001$). Both the workload of AS ($p = .018$) and ST ($p = .03$) were lower than AT. However, there was no difference between AS and ST.

4.3 User Preferences

After each subject completed all the conditions, he/she was asked to answer a survey regarding his/her subjective preferences of each interface. In the survey, the subject

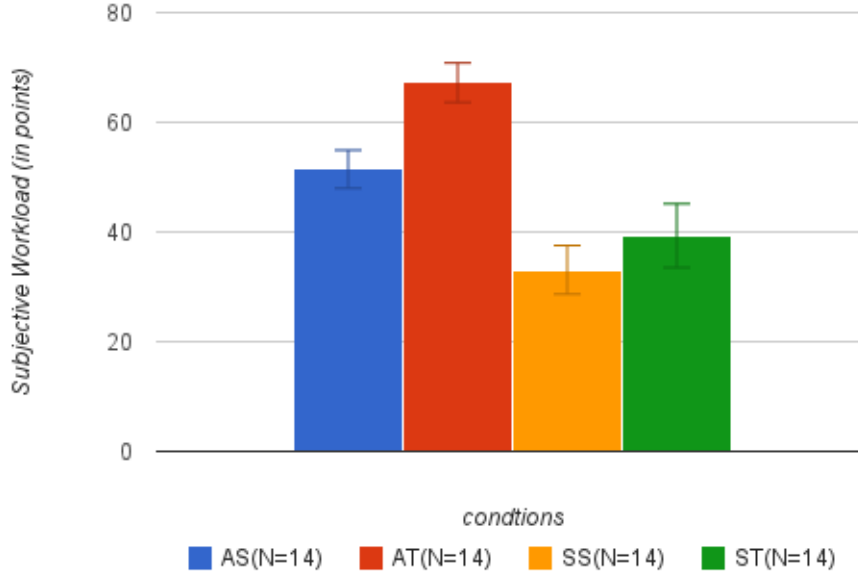


Figure 13: Subjective Workload, N=14

was asked to rank the four interfaces based on overall preference, effectiveness, efficiency, and satisfaction. The result is shown in Table 1. A Wilcoxon Signed Rank Test, the nonparametric equivalent of a paired samples two-tailed t-test, was used for comparing the ranks.

For overall preference, SS was most preferred by the subjects. SS (Md = 1) was ranked significantly higher than AT (Md = 4, $z = -3.508$, $p < .001$). It was also ranked higher than AS (Md = 2.5, $z = -3.040$, $p = .02$) and ST (Md = 2.5, $z = -2.037$, $p = .042$). However, there was no significance between AS and ST.

On effectiveness measure, SS(Md = 1) was ranked higher than AS (Md = 3, $z = -3.01$, $p = .003$) and AT (Md = 4, $z = -3.46$, $p = .001$). ST (Md = 2) was also ranked higher than AS ($z = -2.40$, $p = .016$) and AT ($z = -3.45$, $p = .001$). AS was ranked slightly higher than AT($z = -2.14$, $p = .032$), but the difference between SS and ST was not statistically obvious.

In terms of efficiency, SS(Md = 1) was ranked higher than AS (Md = 3, $z = -3.22$,

Table 1: Median value of each interface with regard to the four aspects (N = 14).

Aspect	AS	AT	SS	ST
Overall Preference	2.5	4	1	2.5
Effectiveness	3	4	1	2
Efficiency	3	4	1	2
Satisfaction	2	4	1	3

$p = .001$), and AT (Md = 4, $z = -3.34$, $p = .001$). ST (Md = 2) was also ranked higher than AS ($z = -2.59$, $p = .01$) and AT ($z = -3.36$, $p = .001$). There was no difference between SS and ST, AS and AT.

With regard to satisfaction, SS (Md = 1) was ranked higher than AS (Md = 2, $z = -2.80$, $p = .005$), ST (Md = 2, $z = -2.49$, $p = .013$), and AT (Md = 4, $z = -3.44$, $p = .001$). AS was ranked higher than AT ($z = -2.58$, $p = .01$). ST was also ranked higher than AT ($z = -3.09$, $p = .02$). AS and ST yielded no obvious significance.

4.4 *Driving Performance*

Subjects' brake response time and the longitudinal variance were measured by the driving simulator. For the brake response time, the average time from each condition of each subject was parsed by a php script. Then a within subjects ANOVA test (two-tailed, confidence interval adjusted by Bonferroni) was used for data analysis. Though there was a little difference between ST (M = 1321.78ms, SD = 47.419) and Baseline Driving (M = 1198.633ms, SD = 48.843), there was no significant variance (shown in Figure 14) found among those five conditions, $F(4,52) = 3.202$, $p > .05$.

For longitudinal performance, the subjects were told to keep 20 meters between the front cars. The data was analyzed in the same way as the brake response time. The result is shown as follows: AS (M = 22.61m, SD = .67), AT (M = 22.58m, SD = .78), Baseline (M = 20.80m, SD = .64), SS (M = 22.23m, SD = .61) and ST (M = 23.00m, SD = .81). A repeated measures ANOVA found no significant effects of

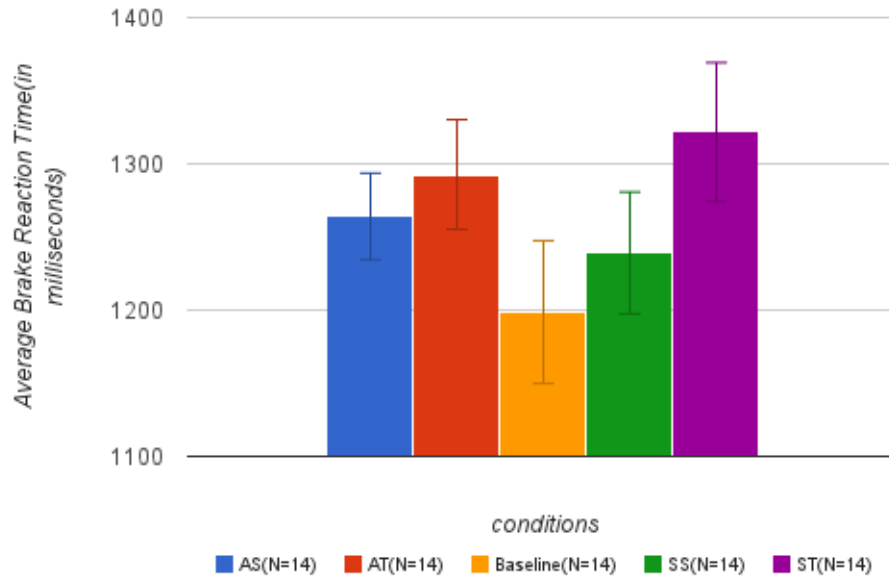


Figure 14: Average Brake Response Time (in milliseconds) N = 14.

condition on average longitudinal distance, $F(4,52) = 2.695$, $p > .05$. (shown in Figure 15)

4.5 *Eye Glance Behavior*

The total frames of subjects' eyeball movements were measured by the remote eye tracker. Then a php script was used to parse each frame. Dwell was defined by the current frame showing 'nothing'. Since the frame rate was 30fps, after all the dwells were recognized, the total dwell duration equaled the number of dwells times 33.3. We calculated the dwell rate by dividing the total dwell time by the average task time that we collected from the secondary task. The long dwell (above 1600ms) was recognized by the one that lasted more than 48 frames. Therefore, the total number of dwells on the informative display, total dwell duration, dwell rate per task, and the number of long dwells were collected. The data was analyzed by a repeated measures ANOVA test.

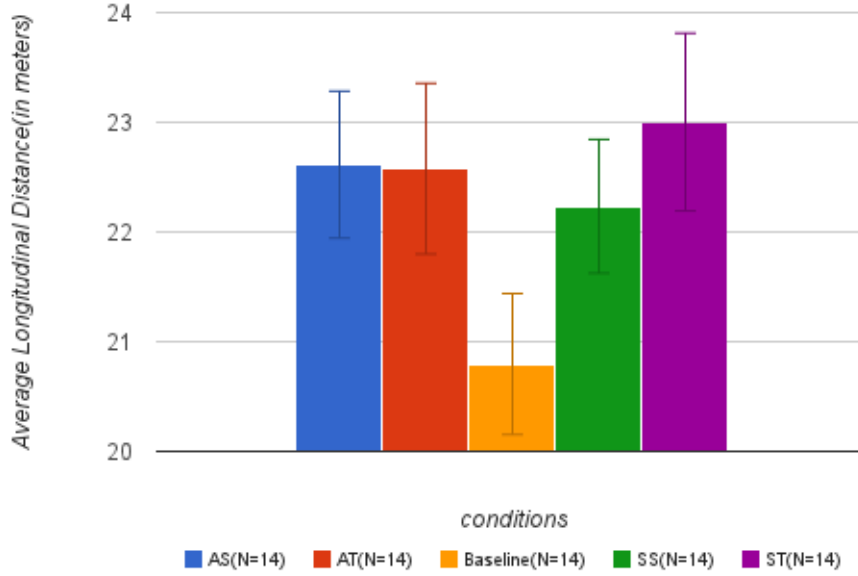


Figure 15: Average Longitudinal Variance (in meters) N = 14.

With regard to total dwell times, five conditions including Baseline were compared. The result showed that the number of dwells in Baseline condition ($M = 36.78s$, $SD = 18.03$) was significantly fewer than that of AT ($p = .005$), and ST ($p = .028$). No difference among other conditions was found. In terms of total dwell duration, AT ($M = 91.40s$, $SD = 15.14$) was found to be significantly longer than Baseline condition ($p = .001$). No significant effects of other conditions on total dwell duration aspect was found. As for dwell rate per task ($F(3,39) = .491$, $p > .05$), four gesture conditions were compared with each other, no significant effect on all the four gesture conditions was found. With regard to the number of long dwells, there was no statistical significance found on the five conditions, $F(4,52) = 2.747$, $p > .05$.

CHAPTER V

DISCUSSION

Among the four interfaces we evaluated in this paper, ST has been widely used in the automobile industry, and SS is also gaining popularity as the multi-touch technology advances. Air gestures, such as AS and AT, have been explored in the research field for a while but have not really been implemented yet. Our study examined ST, SS, AS, and AT in terms of secondary task performance, perceived cognitive workload, driving performance, and eye glance behavior. The result suggested that SS seemed like an ideal way for performing in-car secondary task of music selection. In terms of secondary task efficiency, it was found to be greatly faster than air gestures. It also required the lowest cognitive load, and most preferred by the subjects. Apart from SS, AS and ST were found to be promising competitors for music selection task. Even though there was no difference in driving performance, overall preference and perceived cognitive load. It took longer time for AS to complete the secondary task while ST generated larger number of dwell times.

5.1 Secondary Task Completion Performance

The result indicated that the secondary task completion time of tangible interaction was significantly faster than that of air gestures. One possible reason was that air gestures were currently limited by hand tracking technology. Once a moving hand was detected, it took a few milliseconds for the computer to transfer the data and analyze what gesture it was. During the processing time, the computer was not able to track other hand movement. This caused a minor delay for the air gestures. While for tangible interactions, there was no such problem. Subjects in the study tended to perform gestures really quick on touch screens.

Besides the recognition delay, another possibility was that air gestures always involve a series of hand movements. For example, by observation in the study, we found that when subjects perform the swiping right gesture, some of them naturally begin the movement by weaving their hands towards the left a little bit, then swiping towards the right. If the left movement was clear enough, the hand-tracking device will misrecognize it as a swiping left gesture. This caused the back and forth movement in the music selection task, thus increased the task completion time.

Though there was no significant difference found between SS and ST in our study, one interesting point has been learned by observing the study. When subjects tapped on the virtual buttons quickly, it was difficult for their fingers to stay tapping on the same place, thus leading to 'tapping nothing'.

5.2 Subjective Workload

The perceived workload of each interface was measured. The result showed that SS required the lowest cognitive load. Subjects' familiarity was believed to be a great cause for it. Since subjects knew the screen swiping gesture very well, there was no extra learning curve required in the study. Surprisingly, although 11 out of 14 subjects in this study have never used Hand Tracking device before the study while all of them have used ST in their cars, AS did not require higher cognitive load than ST. It indicated that air swipe gesture is easy to learn, and it might be a suitable way for secondary task in the future.

5.3 User Preferences

SS ranked top in all the four aspects (overall preference, effectiveness, efficiency, satisfaction) we examined. One of our predication is that subjects in the study are generally young; they are familiar and comfortable with the multi-touch interface. Another presumption is that it provides tangible feedback while the targeting area is significantly larger than ST. Therefore it provides them with the best user experience.

Surprisingly, although the effectiveness and efficiency of AS are ranked lower than ST, there is no difference in overall preference and satisfaction aspects. We believe that this resulted from the fact that AS did not fix subjects' hands on a certain target, and that it also generated less dwell count times. Of the four interfaces, AT ranked the lowest in all the aspects. Apart from the limitation for subtle hand gesture recognition, another reason was that air gesture lacked passive feedback. Since there was no additional audio feedback besides the soundtrack name, the subjects in the study tended to feel confused. Two of them claimed that it was difficult for them to tell whether they were tapping in the right or left direction, which caused more error correcting times, thus consuming a large amount of time and effort.

5.4 Driving Performance

The study examined the brake response time and longitudinal variance. There were no significant effects on both aspects found. By examining the mean value of the five conditions, we found a subtle trend that ST has generated longer brake response time and bigger longitudinal variance. More researches are needed in this field to verify this finding.

Further more, according to the other aspects we examined, AT required significant cognitive load and it also generated the longest dwell duration. However, this did not impair the driving performance. We believed the reason was that the driving tasks we assigned to the subjects were too simple that they could handle it with high cognitive load. Therefore, more complex tasks are needed in the future to verify this predication.

5.5 Eye Glance Behavior

The remote eye tracker measured all the eyeball movements of the subjects. In the study, we observed that it is easy that the eye tracking performance was interrupted. Light, subjects' unpredicted head movement, and especially whether the subjects

wear glasses, have been found to affect the tracking accuracy. For AT's total dwell duration time, we observed that subjects tended to look at the head unit once their gesture was not well recognized, or wrongly recognized. AT gesture required subjects to tap their finger in a certain area, we found that it was difficult for subjects to target the area while they are driving. Moreover, limited by the hand recognition technique, and the fact that movement range of AT was not as obvious as AS, it was difficult for the tracking device to recognize, which contributes to the result that AT gained the longest dwell duration time and large number of dwell times. With regard to the large dwell times of ST, we thought that the variance of finger movements resulted in the subjects' using their vision to relocate the virtual buttons. Therefore, it generated more dwell times.

5.6 *Limitation*

Our research suffers from several limitations. First of all, by observation, we found that the size of the subjects' hand affected the accuracy of hand recognition of Leap Motion. If the hand is bigger, it was easier for the device to recognize the gesture, especially for subtle gestures like AT. Secondly, the accuracy of the remote eye tracker in this study was also limited by light and subjects' head movements. We found that if the subject wears a pair of glass, the light of the simulator display will be projected on the glass, thus leads to the wrong tracking or losing track. Moreover, the simulator in this study was a low fidelity one. Some subjects have claimed that the accelerator pedal and brake pedal were difficult to step on. In addition, the driving task for this study was simple; there was no changing lane or turning tasks in the study.

With regard to the secondary task performance, although the variance between air gestures and screen gestures was significant, the amount of task errors remained no difference. However, the error types vary. We observed that some subjects made mistakes by exceeding than what they were asked to, while some made mistakes by

being interfered by the driving tasks (lead car braked). Since we did not record video in the study, the error type was not analyzed.

In terms of user preferences, since our study was limited by the short period of training time, though in this study subjects tended to prefer screen swipe gesture the most, subjects' choices may vary after a long period of time. Therefore, a long-term study in real driving situation may be needed to verify our finding.

CHAPTER VI

CONCLUSION

We compared two air gestures and two tangible interfaces for the task of in-car music selection. Five aspects (driving performance, task completion time, perceived workload, eye glance behavior, and user preferences) of each of the five conditions were measured and analyzed. Our study indicated that the screen swipe gesture for selecting music was efficient, it required the lowest cognitive workload, and was preferred the most by subjects. On the other hand, air swipe gesture, though more time consuming, was preferred by subjects. Moreover, even though there was no cognitive workload variance between air swipe and screen tap gesture, screen tap gesture generated larger amount of dwells. Therefore, air swipe gesture seemed to be a promising competitor for in-car incremental tasks. Air tap gesture in this study was found to be time consuming, required high cognitive load, and ranked the lowest by subjects, so it might not be suitable for in-car use. With respect to driving task, our study revealed no significant difference in driving performance among the four gestures.

As our study suffers from several limitations discussed above, future work may focus on examining the driving performance by providing more complex driving tasks on simulator, and recruiting more participants.

APPENDIX A

SECONDARY TASK PERFORMANCE DATA

Each subject's task completion time and the number of errors were recorded by the application.

Table 2: Secondary Task Completion Time Raw Data(in milliseconds) (subject 01).

Task Index	AS	AT	SS	ST
2	5250	11599	5483	7600
4	9588	9968	7543	7908
6	8611	12261	7705	10654
7	9878	12261	7833	6669
8	13723	12800	8329	10644
9	16353	19862	10783	7249
10	15809	23144	10862	19838
11	19839	24108	7833	7006
12	14718	21353	7217	7459
13	17895	32704	13259	8601
15	18618	24063	12468	14351
18	23627	28346	15175	18995

Table 3: Secondary Task Completion Time Raw Data(in milliseconds) (subject 02).

Task Index	AS	AT	SS	ST
2	5088	6457	6260	7919
4	8086	15095	7230	3554
6	14096	20406	12613	8375
7	21954	9877	4753	9151
8	11521	28131	8623	8938
9	16763	15100	9700	7849
10	20686	12390	5323	10239
11	19570	14838	9990	8245
12	25043	17920	7952	12847
13	29410	16057	12776	7325
15	24718	20620	9897	18773
18	22245	20798	17451	7438

Table 4: Secondary Task Completion Time Raw Data(in milliseconds) (subject 03).

Task Index	AS	AT	SS	ST
2	3961	32961	12029	3471
4	7123	15083	6160	6183
6	8630	19694	6611	7746
7	10532	12441	7275	10978
8	19172	20993	8346	8965
9	33730	19694	7102	7288
10	8630	32515	7753	8014
11	24319	20993	6529	8042
12	15927	26258	5858	8633
13	19965	21287	9682	10458
15	27240	28596	15390	10674
18	35607	32854	12102	10488

Table 5: Secondary Task Completion Time Raw Data(in milliseconds) (subject 04).

Task Index	AS	AT	SS	ST
2	5291	12783	4229	3333
4	10996	8419	4574	6171
6	10953	10900	7485	6939
7	25451	30431	5905	14229
8	10553	9984	6594	7212
9	10112	12413	7222	6695
10	17895	21688	8628	9786
11	11137	17478	9995	5199
12	21690	21920	10256	8968
13	17145	26955	9099	7148
15	15550	18844	16797	14841
18	15167	23169	16128	12666

Table 6: Secondary Task Completion Time Raw Data(in milliseconds) (subject 05).

Task Index	AS	AT	SS	ST
2	10598	4838	3447	7581
4	5362	16313	5102	7047
6	10571	9560	9813	8291
7	10540	6665	7045	6827
8	12027	11290	6586	10066
9	11813	16687	10149	14566
10	12935	33259	10722	12740
11	29651	18000	8400	10727
12	21170	19763	11231	13179
13	27294	38605	7880	15260
15	18879	16556	8893	16490
18	31567	21579	11554	13693

Table 7: Secondary Task Completion Time Raw Data(in milliseconds) (subject 06).

Task Index	AS	AT	SS	ST
2	3889	9280	3822	4284
4	10742	5791	9375	6969
6	15405	9058	10494	9781
7	9821	8641	12488	10173
8	21100	9628	11836	10256
9	17750	10915	14661	11415
10	33281	11048	15104	15468
11	12982	18184	16933	10032
12	15569	12895	20001	15071
13	18599	39990	19510	17240
15	27934	15225	22360	18634
18	45892	20674	24885	19375

Table 8: Secondary Task Completion Time Raw Data(in milliseconds) (subject 07).

Task Index	AS	AT	SS	ST
2	6270	5620	6241	10937
4	9520	7802	9803	10753
6	9238	16114	8171	11735
7	23796	7247	8540	16710
8	8972	21442	7588	13334
9	15957	13197	12675	9049
10	13522	20977	12270	14082
11	13991	13006	7008	5793
12	12165	13496	5111	14179
13	15671	11392	8596	11744
15	17456	13051	10490	21446
18	20462	21553	15672	15274

Table 9: Secondary Task Completion Time Raw Data(in milliseconds) (subject 08).

Task Index	AS	AT	SS	ST
2	9771	14010	10810	30989
4	8238	15722	6500	7309
6	12525	26505	15446	10949
7	19131	18111	7824	7490
8	13824	32768	17687	10472
9	17190	23518	14251	12192
10	15841	35012	12282	11143
11	16263	16945	15721	9221
12	22038	28506	9390	15707
13	24362	22217	13019	13785
15	24203	28071	16868	15186
18	27389	34541	16505	19304

Table 10: Secondary Task Completion Time Raw Data(in milliseconds) (subject 09).

Task Index	AS	AT	SS	ST
2	25832	10229	9844	12592
4	10250	11868	8714	10963
6	12656	24756	13329	11825
7	16766	15361	8970	12816
8	31732	16747	17390	11178
9	19024	17384	13900	13833
10	16841	64493	16012	9389
11	40225	26021	12433	14046
12	25088	23693	54310	24975
13	42745	33214	11884	11387
15	26610	24582	14648	9131
18	36253	38733	14067	13390

Table 11: Secondary Task Completion Time Raw Data(in milliseconds) (subject 10).

Task Index	AS	AT	SS	ST
2	18326	12043	5134	7529
4	4301	9300	4589	13395
6	9650	15780	9391	10100
7	7767	15050	5594	12119
8	10853	42332	8141	8686
9	14215	14748	8048	11484
10	15489	41815	5521	13216
11	19553	139617	18503	11143
12	9973	28940	10291	11254
13	12400	29590	6692	10020
15	16052	62683	6141	14333
18	23038	60523	15223	15648

Table 12: Secondary Task Completion Time Raw Data(in milliseconds) (subject 11).

Task Index	AS	AT	SS	ST
2	4748	7968	8386	5400
4	10189	11608	12494	7528
6	16674	18335	10530	8345
7	20455	11754	12457	8905
8	25235	36782	10984	15218
9	13653	14748	13167	11532
10	16855	20011	14881	15118
11	34705	14336	12251	14733
12	19390	20703	12274	15335
13	49509	22481	22836	16432
15	25986	25326	19235	28778
18	32375	27026	25500	19237

Table 13: Secondary Task Completion Time Raw Data(in milliseconds) (subject 12).

Task Index	AS	AT	SS	ST
2	4762	7556	6187	3018
4	8149	8012	9891	16182
6	14836	15868	9013	5354
7	10395	10526	7309	16102
8	16379	12749	17623	10256
9	10072	23996	8426	11299
10	24722	19952	10468	11024
11	20635	25981	22139	10156
12	19418	28111	13104	15475
13	15596	30424	9081	13167
15	46461	21241	8547	20148
18	25532	47838	15983	16083

Table 14: Secondary Task Completion Time Raw Data(in milliseconds) (subject 13).

Task Index	AS	AT	SS	ST
2	21403	3522	7259	3547
4	6195	7276	7746	7585
6	11246	10384	9998	13168
7	6737	8253	17626	10781
8	7159	16905	9208	7287
9	11119	14575	4427	7092
10	15425	20597	5891	12787
11	44585	25302	11490	5093
12	13495	17280	10307	12971
13	17585	6084	10927	12154
15	20064	21851	10278	9651
18	21693	32143	11142	12240

Table 15: Secondary Task Completion Time Raw Data(in milliseconds) (subject 14).

Task Index	AS	AT	SS	ST
2	4073	24438	3784	12748
4	8605	7788	6750	7127
6	8838	11980	10794	10770
7	13734	9863	9569	11378
8	13130	15973	12888	10850
9	12154	12608	13664	11095
10	15688	17911	12109	16406
11	17135	18625	17255	13697
12	16313	15721	14149	14494
13	17565	26633	13398	15057
15	19550	24613	16438	12627
18	22816	24059	20249	15960

Table 16: Average Secondary Task Completion Time(in milliseconds), N = 14.

Subject	AS(N = 14)	AT(N = 14)	SS(N = 14)	ST(N = 14)
01	14492.41667	19372.41667	9540.833333	10581.16667
02	18265	16474.08333	9380.666667	9221.083333
03	17903	23614.08333	8736.416667	8411.666667
04	14328.33333	17915.33333	8909.333333	8598.916667
05	16867.25	17759.58333	8401.833333	11372.25
06	19413.66667	14277.41667	15122.41667	12391.5
07	13918.33333	13741.41667	9347.083333	12919.66667
08	17564.58333	24660.5	13025.25	13645.58333
09	25335.16667	25590.08333	16291.75	12960.41667
10	13468.08333	39882.25	8605.666667	11577.25
11	22481.16667	19256.5	14582.91667	13880.08333
12	18079.75	21021.16667	11480.91667	12355.33333
13	16392.16667	15347.66667	9691.583333	9529.666667
14	14133.41667	17517.66667	12587.25	12684.08333

Table 17: Number of Secondary Task Errors(in times),N = 14.

Subject	AS(N = 14)	AT(N = 14)	SS(N = 14)	ST(N = 14)
01	0	0	0	0
02	1	1	0	0
03	0	1	0	0
04	0	1	0	0
05	0	0	0	0
06	0	0	0	0
07	0	0	2	0
08	0	0	0	0
09	1	2	0	0
10	0	4	0	0
11	0	0	0	0
12	2	0	0	0
13	0	0	0	0
14	0	1	0	0

APPENDIX B

NASA TLX WORKLOAD DATA

The NASA TLX survey examined six aspects of a task, including mental demand, physical demand, temporal demand, performance, effort, and frustration, by letting the subjects compare each pair of aspects based on their subjective opinions. The output of the survey was a number ranging from 0 to 100; it indicated the overall workload score. The higher the score was, the higher the workload the subject perceived.

Appendix C.

Subject ID: _____ Task ID: _____

RATING SHEET

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

EFFORT

Low High

FRUSTRATION

Low High

16

17

Table 18: Overall Workload Score(in points), N = 14.

Subject	AS(N = 14)	AT(N = 14)	SS(N = 14)	ST(N = 14)
01	43.33	63.14	31.47	65.66
02	73.32	67.67	19.94	9.14
03	39.33	62.34	2.27	9.34
04	41.59	66.06	19.47	13.53
05	60.47	67.01	30.92	35.8
06	51.4	54.66	47.07	70.19
07	30.13	70.21	13.67	14.21
08	38.4	89.06	51.47	46.4
09	55.53	58.47	47.66	38.74
10	76.41	99.27	23.86	66.73
11	49.55	43.8	28.74	30.33
12	56.4	66.4	35.66	60.73
13	56.13	64.53	52.06	43.8
14	48.2	68.47	59.06	45.93

APPENDIX C

USER PREFERENCES DATA

Subjects in the questionnaire were asked to rank the four types of interfaces based on their perceived overall preference, effectiveness, efficiency, and satisfaction. Number 1 stands for most preferred, number 4 stands for least preferred.

Comparing cognitive loads of gestures on drivers

1. Please rank the four interfaces based on **Overall Preference** (1 – most preferred, 4 – least preferred).

☐ Air Swipe
☐ Air Tap
☐ Screen Swipe
☐ Screen Tap

2. Please rank the four interfaces based on **Effectiveness** (1 – most preferred, 4 – least preferred).

☐ Air Swipe
☐ Air Tap
☐ Screen Swipe
☐ Screen Tap

3. Please rank the four interfaces based on **Efficiency** (1 – most preferred, 4 – least preferred).

☐ Air Swipe
☐ Air Tap
☐ Screen Swipe
☐ Screen Tap

4. Please rank the four interfaces based on **Satisfaction** (1 – most preferred, 4 – least preferred).

☐ Air Swipe
☐ Air Tap
☐ Screen Swipe
☐ Screen Tap

Table 19: Overall Preference (N = 14).

Subject	AS(N = 14)	AT(N = 14)	SS(N = 14)	ST(N = 14)
01	2	4	1	3
02	4	3	2	1
03	2	4	1	3
04	3	4	1	2
05	3	2	1	4
06	3	4	1	2
07	2	4	1	3
08	1	4	2	3
09	4	3	2	1
10	2	4	1	3
11	3	4	1	2
12	3	4	1	2
13	2	4	3	1
14	2	4	1	3

Table 20: Effectiveness(N = 14).

Subject	AS(N = 14)	AT(N = 14)	SS(N = 14)	ST(N = 14)
01	3	4	1	2
02	4	3	2	1
03	3	4	2	1
04	3	4	1	2
05	4	3	1	2
06	3	4	1	2
07	3	4	1	2
08	1	4	2	3
09	2	4	1	3
10	2	4	3	1
11	2	4	1	3
12	3	4	2	1
13	3	4	2	1
14	4	3	1	2

Table 21: Efficiency(N = 14).

Subject	AS(N = 14)	AT(N = 14)	SS(N = 14)	ST(N = 14)
01	2	4	1	3
02	4	3	2	1
03	3	4	2	1
04	3	4	1	2
05	4	3	1	2
06	3	4	1	2
07	3	4	1	2
08	2	4	1	3
09	4	3	1	2
10	3	4	3	2
11	2	4	1	3
12	4	3	2	1
13	3	4	2	1
14	4	3	1	2

Table 22: Satisfaction(N = 14).

Subject	AS(N = 14)	AT(N = 14)	SS(N = 14)	ST(N = 14)
01	1	4	2	3
02	4	3	2	1
03	3	4	1	2
04	3	4	1	2
05	4	2	1	3
06	3	4	1	2
07	2	4	1	3
08	2	4	1	3
09	2	4	1	3
10	2	4	1	3
11	2	4	1	3
12	3	4	1	2
13	2	4	3	1
14	2	4	1	3

APPENDIX D

DRIVING PERFORMANCE DATA

Table 23: Average Brake Response Time (in milliseconds), N = 14.

Subject	AS(N = 14)	AT(N = 14)	Baseline(N = 14)	SS(N = 14)	ST(N = 14)
01	1176.619048	1160	1355.230769	1118.941176	1283.625
02	1217.090909	1338.619048	1076.470588	1136.882353	1198.933333
03	1279.84	1251.206897	1100.625	1291.0625	1210.333333
04	1388.941176	1363.88	1221	1438.166667	1440.263158
05	1208.478261	1195.423077	1028.071429	1078.8	1065.272727
06	1248.75	1105.272727	1235.625	1234.625	1177.772727
07	1141.619048	1147.105263	1003.882353	1130	1466.6875
08	1189.774194	1182.548387	1097.833333	1052.772727	1133.083333
09	1471.709677	1548	1554.785714	1569.26087	1539.076923
10	1250.25	1450.232558	1084.8125	1108.9	1259.47619
11	1230.384615	1417.217391	1527.6	1454.47619	1698.380952
12	1282.136364	1434.909091	1081.125	1226.611111	1448.0625
13	1481.125	1357.277778	1355.428571	1292.076923	1370.846154
14	1130.444444	1146.041667	1058.375	1214.666667	1213.1

Table 24: Average Longitudinal Variance (in meters), N = 14.

Subject	AS(N = 14)	AT(N = 14)	Baseline(N = 14)	SS(N = 14)	ST(N = 14)
01	21.75047619	21.204656	22.50713009	22.51040046	22.73112829
02	25.11724099	22.04415223	18.53202714	19.97476567	23.90972667
03	23.4278843	22.18114391	21.00176861	22.7801054	24.53162679
04	25.46581368	21.63044762	18.23595934	20.31328478	21.24973289
05	21.50570189	21.35566282	24.81917301	19.09433879	18.33770576
06	19.91074994	19.68758836	19.48335569	21.1235291	19.93306948
07	22.72664758	26.39111467	18.98720713	21.56203208	26.77885085
08	18.38290499	20.86388576	16.53042598	22.56957806	17.9605614
09	20.95431625	25.84797622	23.83332508	22.4269292	27.06671466
10	19.55921035	16.20380359	19.37444452	20.46435926	21.4392564
11	22.22840382	22.55005135	21.56788478	21.64964558	21.36014169
12	23.59983841	23.70305309	20.96543244	23.23643654	26.25081148
13	27.30225657	26.73607235	23.76871882	27.44849181	25.48992629
14	24.6709509	25.68017532	21.54115611	26.12254029	24.98648138

APPENDIX E

EYE GLANCE BEHAVIOR DATA

Table 25: Total Eye Dwell Duration (in milliseconds), N = 14.

Subject	AS(N = 14)	AT(N = 14)	Baseline(N = 14)	SS(N = 14)	ST(N = 14)
01	672634.5	183303.32	139894.64	222844.56	303827.42
02	44808.96	62479.16	31739.68	26872.04	112755.88
03	152463.82	126125.22	38374.34	217243.44	53777.42
04	49976.66	101553.64	5934.52	62179.1	79282.52
05	110522.1	154564.24	11935.72	23371.34	43708.74
06	63446.02	26038.54	3200.64	32839.9	15169.7
07	102153.76	75681.8	3834.1	25671.8	23904.78
08	2567.18	16336.6	7768.22	14369.54	10102.02
09	78382.34	61745.68	1700.34	29139.16	36473.96
10	60845.5	34773.62	14402.88	260985.52	64079.48
11	57011.4	171834.36	113922.78	82983.26	71180.9
12	51943.72	139061.14	155964.52	26938.72	98152.96
13	13869.44	101787.02	800.16	8435.02	62979.26
14	76548.64	71881.04	12435.82	75748.48	77515.5

Table 26: Total Number of Dwells (in times), N=14.

Subject	AS(N = 14)	AT(N = 14)	Baseline(N = 14)	SS(N = 14)	ST(N = 14)
01	630	239	176	232	394
02	60	92	51	39	97
03	230	199	53	314	100
04	55	62	7	46	55
05	104	121	10	22	47
06	67	38	4	50	19
07	117	100	4	26	25
08	3	19	11	17	13
09	125	109	0	40	70
10	105	45	24	335	114
11	56	29	20	22	44
12	91	235	25	41	160
13	22	94	0	6	59
14	122	107	12	86	89

Table 27: Total Number of Long Dwells (in times) ,N = 14.

Subject	AS(N=14)	AT(N=14)	Baseline(N=14)	SS(N=14)	ST(N=14)
01	30	21	19	37	34
02	3	4	0	1	16
03	12	10	2	19	8
04	6	26	2	19	20
05	18	27	1	4	7
06	9	0	0	1	1
07	10	8	1	2	4
08	0	2	0	1	0
09	14	28	0	35	106
10	2	0	0	18	3
11	5	11	12	6	11
12	0	3	9	0	5
13	0	22	0	1	12
14	3	2	2	7	7

Table 28: The Dwell Rate Per Task, N = 14.

Subject	AS(N = 14)	AT(N = 14)	SS(N = 14)	ST(N = 14)
01	29.91991074	9.51903617	15.28120643	21.88945215
02	2.478405951	2.972202304	2.340583142	9.126089678
03	9.301016949	8.217875899	22.41568096	5.643158557
04	3.536063585	5.797212719	4.939847862	6.250551807
05	6.051032028	9.382266489	2.491436998	4.7400873
06	0.982766417	0.39875333	0.639241585	11.97702909
07	5.705957661	3.204943378	2.938481643	2.841860115
08	0.179168082	0.911878093	1.612863664	1.174801477
09	4.647013591	3.476752739	3.468190673	3.207277364
10	3.134158067	2.435568059	17.25818867	5.171244805
11	4.09613699	12.50485042	8.877984398	5.509499729
12	2.957298731	5.63902354	2.068192165	7.193020452
13	0.547438277	3.977596269	0.51774794	4.859354573
14	5.281978966	3.710483892	7.939398725	7.325798982

REFERENCES

- [1] Christiansen, L. H., Frederiksen, N. Y., Jensen, B. S., Ranch, A., Skov, M. B., & Thiruravichandran, N. (2011). Don't look at me, i'm talking to you: investigating input and output modalities for in-vehicle systems. In *Human-Computer Interaction—INTERACT 2011* (pp. 675-691). Springer Berlin Heidelberg.
- [2] Pickering, C. A., Burnham, K. J., & Richardson, M. J. (2007, June). A research study of hand gesture recognition technologies and applications for human vehicle interaction. In *3rd Conf. on Automotive Electronics*.
- [3] J  ger, M. G., Skov, M. B., & Thomassen, N. G. (2008, April). You can touch, but you can't look: interacting with in-vehicle systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1139-1148). ACM.
- [4] Akyol, S., Canzler, U., Bengler, K., & Hahn, W. (2000, November). Gesture Control for Use in Automobiles. In *MVA* (pp. 349-352).
- [5] Alpern, M., & Minardo, K. (2003, April). Developing a car gesture interface for use as a secondary task. In *CHI'03 extended abstracts on Human factors in computing systems* (pp. 932-933). ACM.
- [6] Kujala, T. (2013). Browsing the information highway while driving: three in-vehicle touch screen scrolling methods and driver distraction. *Personal and ubiquitous computing*, 17(5), 815-823.
- [7] Lasch, A., & Kujala, T. (2012, October). Designing browsing for in-car music player: effects of touch screen scrolling techniques, items per page and screen orientation on driver distraction. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 41-48). ACM.
- [8] Leap Motion.(2014, August). From [http //en.wikipedia.org/wiki/Leap_Motion](http://en.wikipedia.org/wiki/Leap_Motion).
- [9] Jeon, M., Walker, B. N., & Srivastava, A. (2012). 'Spindex'(Speech Index) Enhances Menus on Touch Screen Devices with Tapping, Wheeling, and Flicking. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19(2), 14.
- [10] Damon Lavrinc.(2014, July). Gesture Controls Are Coming To New Cars Next Year. From: <http://jalopnik.com/gesture-controls-are-coming-to-new-cars-next-year-1598692093>
- [11] Kamp, J.F., Poirier, F., Doignon, P.H. (1999). 'Interaction with In-Vehicle Systems', Paper. Vannes, France: Universite De Bretagne-Sud.

- [12] Engstr m, J., Johansson, E., &  stlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2), 97-120.
- [13] Muller, C., Weinberg, G., & Vetro, A. (2011). Multimodal input in the car, today and tomorrow. *IEEE MultiMedia*, 18(1), 98-103.
- [14] Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(4), 631-640.
- [15] Lee, S., & Hong, S. H. (2005). Design of an integrated wearable multimedia interface for in-vehicle telematics. In *Advances in Multimedia Information Processing-PCM 2004* (pp. 113-120). Springer Berlin Heidelberg.
- [16] Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J. D., ... & Knipling, R. R. (2006). The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment (No. HS-810 593).
- [17] Stutts, J. C. (2005, October). How risky is distracted driving?'What crash data reveals'. In *International Conference on Distracted Driving*. Toronto, Canada.
- [18] Palinko, O., Kun, A. L., Shyrov, A., & Heeman, P. (2010, March). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 141-144). ACM.
- [19] Lyons, K., Brashear, H., Westeyn, T., Kim, J. S., & Starner, T. (2007). Gart: The gesture and activity recognition toolkit. In *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments* (pp. 718-727). Springer Berlin Heidelberg.
- [20] Subaru USA 2013 Subaru BRZ Limited Stereo Head Unit with 6.1-Inch Touch Screen From: <http://www.tastyautos.com/2013-subaru-brz-limited-stereo-head-unit-with-6-1-inch-touch-screen>
- [21] Norman, D. A. (2002). *The design of everyday things*. Basic books.
- [22] AAA Foundation for Traffic Safety, (2014) From <http://newsroom.aaa.com/2014/10/imperfect-hands-free-systems-causing-potentially-unsafe-driver-distractions/>
- [23] de Winter, J., Van Leeuwen, P. M., & Happee, R. (2012, August). Advantages and Disadvantages of Driving Simulators: A Discussion. In *Proceedings of Measuring Behavior* (pp. 47-50).

- [24] BÃrdard, M., Parkkari, M., Weaver, B., Riendeau, J., & Dahlquist, M. (2010). Assessment of driving performance using a simulator protocol: Validity and reproducibility. *American Journal of Occupational Therapy*, 64(2), 336-340.
- [25] Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 26(3), 257-268.
- [26] Campbell, J. L., Carney, C., & Kantowitz, B. H. (1998). Human factors design guidelines for advanced traveler information systems (ATIS) and commercial vehicle operations (CVO) (No. FHWA-RD-98-057,).
- [27] Antin, J. F., Dingus, T. A., Hulse, M. C., & Wierwille, W. W. (1990). An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. *International Journal of Man-Machine Studies*, 33(5), 581-594.
- [28] Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(1), 196-205.
- [29] Billingham, M., & Buxton, B. (2011). Gesture based interaction. *Haptic input*, 24.
- [30] May, K. R., Gable, T. M., & Walker, B. N. (2014). Taxonomic Description of One-Handed Air Gesture Preferences for In-Vehicle Menu Control.
- [31] Tsimhoni, O., & Green, P. (2001, October). Visual demand of driving and the execution of display-intensive in-vehicle tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 45, No. 23, pp. 1586-1590). SAGE Publications.
- [32] Gellatly, A. W. (1997). The use of speech recognition technology in automotive applications (Doctoral dissertation, Virginia Polytechnic Institute and State University).
- [33] Bellotti, F., De Gloria, A., Montanari, R., Dosio, N., & Morreale, D. (2005). COMUNICAR: designing a multimedia, context-aware human-machine interface for cars. *Cognition, Technology & Work*, 7(1), 36-45.
- [34] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.
- [35] Noy, Y. I., Lemoine, T. L., Klachan, C., & Burns, P. C. (2004). Task interruptability and duration as measures of visual distraction. *Applied Ergonomics*, 35(3), 207-213.